DOCUMENT RESUME

ED 093 997 TM 003 854

AUTHOR Novak, Carl D.
TITLE An Empirical Investigation of Multiple Matrix Sampling in an Elementary School Setting.
PUB DATE Apr 74
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS Academic Achievement; *Achievement Tests; Analysis of Variance; *Comparative Analysis; *Elementary Schools; *Item Sampling; Standardized Tests; Tests
IDENTIFIERS Iowa Tests of Basic Skills; *Lincoln Nebraska Public Schools

ABSTRACT

This study involving 352 students was designed to verify empirically the a priori use of multiple matrix sampling procedures in an elementary school using a nationally normed, commercizlly published achievement test. The study focused on effect of changes in item context, effect of previous exposure to items, and relative effectiveness of multiple matrix sampling procedures. Results indicated that multiple matrix sampling estimates of the mean were more accurate and estimates of the variance were as accurate as comparable examinee sampling estimates. Changes in item context affected matrix sample variance estimates. Previous exposure to items affected matrix sample mean estimates. (Author)

[1]An Empirical Investigation of Multiple Matrix Sampling in an Elementary School Setting

by

Carl D. Novak

Lincoln Public Schools
Lincoln, Nebraska

Introduction

Multiple matrix sampling is a psychometric procedure for estimating group parameters. It involves the simultaneous random sampling of both items and examinees. Although educational research specialists and educators in general have long been aware of the advantages of examinee sampling, it was not until the sixties that research specialists began experimenting with matrix sampling procedures.

The early studies helped clarify the potential of the item sampling option of multiple matrix sampling. Many of them, however, followed a research paradigm that restricted the generalizability of the findings. The paradigm involved the extraction of multiple matrix sampling estimates from an existing matrix of examinee-item responses collected by the administration of an entire test to a group of students. In direct contrast to sampling from an existing response matrix is the way in which matrix sampling would be used in an applied situation; each student would take only a fractional sample of the items on the test. The differences between having a student respond to all 100 items on a test and having him respond to, for example, five items is obvious. Any of a number of error factors, such as anxiety, motivation, fatigue, etc., could operate to make examinees respond differently

[1]A paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, April 17, 1974.

in the two situations.

A review of the literature identified several studies that deviated from the early paradigm and that were relevant to the issues being investigated in this study. Owens and Stufflebeam (1967) administered matrix tests to 3330 fourth grade students from both advantaged and disadvantaged neighborhoods. Each student responded to a matrix test of either three, six, or 12 items. The authors concluded that the item sample estimates of the mean were generally closer to the computed population value than comparable examinee sample estimates. Item sample estimates of the variance were not as precise as variance estimates of comparable examinee sample estimates. Although the students in the Owens and Stufflebeam study responded to matrix tests, the remaining items of the test being sampled, the Metropolitan Reading Test, were administered in conjunction with the administration of the matrix tests.

Cahen, Romberg, and Zwirner collaborated on a pair of studies that involved the administration of matrix tests (1970, 1973). The first study involved the use of multiple matrix sampling to estimate the performance of ninth grade students from 81 schools on a 50-item mathematics test. The matrix sample estimates of the mean preserved the relative rank ordering of the schools; however, the multiple matrix estimates were systematically higher than the actual school means.

The second Cahen, Romberg, and Zwirner study included an interesting variation. The test sample was a 24-item Project Talent Mathematics Test. The population of interest was twelfth grade students from 35 schools that participated in the National Longitudinal Study of Mathematical Abilities. Half of the students in each school took the total test on the first day of testing and the item sample subtests on the second day of testing. The other half took only the item sample subtest on the second day. The authors concluded that matrix sample estimates again provided reasonable estimates of the group mean and that taking the total test on the first day did not affect student performance on the matrix tests on the second day of testing.

Neither the Owens and Stufflebeam study nor the two studies by Cahen, Romberg, and Zwirner were specifically designed to test for the existence of a context effect. Sirotnik (1970) designed a study to test directly for context effect. The Sirotnik study involved the direct comparison of the matrix sample estimates extracted from total test data (treatment A) with matrix sample estimates computed from data collected by the independent administration of matrix tests (treatment B). Matrix sample estimates of student performance on three different tests, vocabulary, mathematics, and attitude toward reading, were collected under each treatment. A multivariate analysis of variance design was utilized to test for systematic differences due to context. None were found. Sirotnik pointed out in his discussion that an insignificant result on a single test of a null hypothesis does not prove that the null hypothesis is true and indicated a need for the study to be replicated.

Statement of the Problem

The purpose of this study was to test the feasibility of the a priori use of multiple matrix sampling procedures in a particular setting, the elementary school, with a particular type of instrument, a commonly used, nationally normed, commercially published achievement test. Specifically, the study focused on three hypotheses:

1. The change in item context which is necessitated by the a priori or applied use of multiple matrix sampling does not significantly affect the matrix sample estimates of the population mean and variance.

2. Recent previous exposure to the items being sampled does not significantly affect the matrix sample estimates of the population mean and variance.

3. The a priori use of multiple matrix sampling procedures described in this study will result in estimates of the population mean and variance that are as accurate as the estimates obtained from examinee sampling procedures based on the same number of observations.

Methodology

The study involved 124 fourth grade students, 119 fifth grade students, and 109 sixth grade students who were attending two different elementary schools. Both elementary schools were part of a consolidated Nebraska school district.

The instrument. The tests used in the study were three subtests of Form 5 of the *Iowa Tests of Basic Skills*. The criteria for the selection of the subtests were that (1) each subtest be representative of a different content area, (2) each subtest use different item formats, (3) each subtest lend itself to simple sampling procedures, and (4) each subtest contain items reproducible in black and white offset. These criteria eliminated most of the other subtests. The three subtests chosen, Vocabulary, Spelling, and Mathematics Concepts, were representative of three different item formats and two distinctly different content areas.

Sampling plan. Each of the nine subtests, three subtests at three grade levels, was subdivided into six matrix tests. A stratified sampling plan was used to assign each item within each subtest (for each grade) to one of the matrix tests. The items were stratified according to difficulty. The stratified sampling plan was used to insure that the matrix tests were of approximately equal difficulty levels. The decision to use *six* matrix tests per subtest was based on the need to have the matrix tests large enough so that examinees would see them as having substance but yet have the tests short enough so that the use of multiple matrix sampling resulted in a viable savings of time. The number of items within any individual matrix test ranged from six to eight. The number of items within any set of matrix test consisting of a vocabulary test, a spelling test and a mathematics test varied from 18-20 for fourth graders, 21-23 for fifth graders and 21-24 for sixth graders.

The matrix tests were randomly assigned to examinees. Matrix tests for each subtest were assigned independently so that most examinees were assigned unique combinations of matrix tests. Two sets of matrix tests were assigned to each participant. The first set was used to collect data for Estimates 1 and 2; the second was used to collect data for Estimates 3 and 4.

Procedures. Three sets of data were collected, the results of the two administrations of the matrix tests and the administration of the *Iowa Tests of Basic Skills* battery. The three sets of data represented four unique combinations of context and exposure. A set of nine multiple matrix sample estimates, one for each of the three subtests at each of the three grade levels, was computed for data representing each of the four combinations of context and exposure. The following four sets of multiple matrix sample estimates are summarized in Table I.

Estimate 1. Data were collected by the administration of Set 1 of the matrix sample tests (matrix context), and the examinees had not previously responded to the items (no exposure).

Estimate 2. Data were collected during the administration of the *Iowa Tests of Basic Skills* battery (normal context); and the examinees had, as a result of Estimate 1, previously responded to the items in their matrix tests (previous exposure).

Estimate 3. Data were collected during the administration of the *Iowa Tests of Basic Skills* battery (normal context); however, since the second assignment of matrix tests was used, the examinees had not previously responded to the items (no exposure).

Estimate 4. Data were collected by the administration of Set 2 of the matrix sample tests (matrix context); and the examinees had, as a result of the administration of the entire *Iowa Tests of Basic Skills*

TABLE I

CONDITIONS OF CONTEXT AND EXPOSURE FOR THE FOUR SETS OF MATRIX SAMPLE ESTIMATES

Estimate	Context	Exposure
1	Derived from the administration of Set 1 of the matrix sample tests	Students had not previously been exposed to any of the items included in the matrix tests
2	Derived post hoc from data collected during the administration of the entire ITBS* battery simulating administration of Set 1 of the matrix sample tests	Students had previous exposure to the items sampled during the administration of the matrix tests associated with Estimate 1
3	Derived post hoc from data collected during the administration of the entire ITBS battery simulating administration of Set 2 of the matrix sample tests	Individual students had not previously been exposed to any of the items included in the Estimate 3 sample
4	Derived from administration of Set 2 of the matrix sample tests	Students had previously been exposed to all items during the administration of the entire ITBS battery

*Iowa Tests of Basic Skills

battery, previously responded to the items in their matrix tests (previous exposure).

In addition to the four sets of multiple matrix sample estimates, ten sets of examinee sample estimates were computed. The examinee sample estimates were equivalent to the matrix sample estimates in that both were based on the same number of examinee-item responses. The examinee sample estimates were computed by randomly selecting 21 fifth graders, 20 fourth graders, and 18 sixth graders. The random selection was replicated 30 times since one replication was necessary to estimate each of the three Iowa Tests of Basic Skills subtests for each of the 10 sets of estimates. The number of observations used in the matrix sample estimates and the examinee sample estimates are summarized in Table II.

Analysis

The analysis consisted of comparing the multiple matrix estimates with the population parameters, the matrix context multiple matrix sample estimates with the post hoc matrix estimates, the previous exposure matrix sample estimates with the no previous exposure estimates, and the a priori multiple matrix sample estimates (Estimate 1) with the examinee sample estimate.

All estimates, whether matrix or examinee sampling estimates, were at one point or other compared with the counterpart population parameters. For the comparison to be valid, the population parameters must be valid. If the prior administration of the matrix sample tests biased the population parameters, then an adjustment would have to be made to compensate for the bias.

The most logical effect of the prior administration of the matrix tests was higher scores on the subsequent testings. Such an

TABLE II

NUMBER OF OBSERVATIONS (ITEMS x EXAMINEES) INCLUDED IN EACH OF THE IOWA TESTS OF BASIC SKILLS SUBTESTS, THE MATRIX TESTS, AND THE EXAMINEE SAMPLES

ITBS* Subtest	Total Test	Matrix Test			Examinee Sample	
	Total Number of Observations	Number of Observations Set 1	Number of Observations Set 2	Percent of Total Observations	Number of Observations	Percent of Total Observations
Grade 4					N = 21	
Vocabulary	4712	788	784	16.7	798	16.9
Spelling	4712	788	786	16.7	798	16.9
Mathematics	4464	744	742	16.6	756	16.9
Grade 5					N = 20	
Vocabulary	5117	853	854	16.7	860	16.8
Spelling	5117	852	855	16.7	860	16.8
Mathematics	4993	833	833	16.7	840	16.8
Grade 6					N = 18	
Vocabulary	5014	835	835	16.7	828	16.5
Spelling	5014	835	835	16.7	828	16.5
Mathematics	4905	820	818	16.7	810	16.5

*Iowa Tests of Basic Skills

effect should be most noticeable in matrix Estimate 2, which was based on the same items that were administered in the a priori matrix sampling. There were no significant differences between matrix sampling Estimate 2 and either Estimate 1 or Estimate 3. In fact, the estimates in set 2 tended to be slightly lower than the estimates in the other sets. Therefore, no adjustment in the population parameter was considered to be necessary. The matrix sample estimates of the means and variances are found in Tables III and IV.

Analysis of Context and Exposure Effects

A 2 x 2 x 3 multivariate analysis of variance design was utilized to test for the existence of a context and/or exposure effect. The dependent variables were the individual matrix test estimates for each of the three subtests, Vocabulary, Spelling and Mathematics. Two separate analyses were run. In the first, the mean scores were used as criterion variables, while in the second, the variance scores were used as criterion variables. Tests of significance were computed for three main effects, context, exposure and grade level, and for the following interactions: context-exposure, context-grade, exposure-grade, and context-exposure-grade. Significant F ratios were found for context effect utilizing variances as the criterion (F statistic of 13.76 for 3, 13 df, p of less than 0.01) and for exposure effect utilizing the means as the criterion (F statistics of 17.73 for 3, 13 df, p of less than 0.01). None of the interactions were significant at the .05 level. The Summary Tables for the tests of main effects are presented in Table V (A).

TABLE III

MULTIPLE MATRIX SAMPLE ESTIMATES OF THE MEAN

ITBS* Subtest	Population Mean	Estimate 1	Estimate 2	Estimate 3	Estimate 4
Grade 4					
Vocabulary	21.177	20.595	22.046	21.278**	22.199
Spelling	18.798	19.289**	16.975	18.807***	17.791**
Mathematics	17.935	18.242***	18.290***	18.387***	19.597
Grade 5					
Vocabulary	25.202	24.852**	26.869	25.377**	27.139
Spelling	19.807	20.844**	17.664	20.117**	19.614***
Mathematics	19.933	19.866***	20.067**	18.908	21.479
Grade 6					
Vocabulary	25.257	24.625**	26.168**	25.617**	26.278**
Spelling	20.046	20.714**	18.290	20.493**	19.392**
Mathematics	19.450	18.493	20.579	19.199**	21.620

*Iowa Tests of Basic Skills

**Closer to the mean than five of the ten randomly drawn equivalent examinee sample estimates

***Closer to the mean than all ten randomly drawn equivalent examinee sample estimates

TABLE IV

MULTIPLE MATRIX SAMPLE ESTIMATES OF THE VARIANCE

ITBS* Subtest	Population Variance	Estimate 1	Estimate 2	Estimate 3	Estimate 4
Grade 4					
Vocabulary	53.074	52.704***	54.103***	45.767**	59.719**
Spelling	59.740	20.184	45.864	76.972	40.482
Mathematics	34.825	30.397**	33.552**	30.471**	31.712**
Grade 5					
Vocabulary	59.739	59.238***	52.148**	61.556***	72.999
Spelling	66.784	55.599**	53.08[illegible]	56.370**	46.108
Mathematics	37.029	39.252***	43.148**	39.566***	34.665***
Grade 6					
Vocabulary	66.711	65.279**	59.417**	62.666**	89.471**
Spelling	62.896	56.493**	74.544**	66.664**	37.712
Mathematics	44.879	60.923	48.153***	38.646**	57.455

*Iowa Tests of Basic Skills

**Closer to the mean than five of the ten randomly drawn equivalent examinee sample estimates

***Closer to the mean than all ten randomly drawn equivalent examinee sample estimates

TABLE V (A)

SUMMARY TABLES FOR TESTS OF MAIN EFFECTS

SUMMARY TABLE FOR HYPOTHESIS 1; CONTEXT EFFECT ON THE ESTIMATES OF THE MEAN*

Computation	Vocabulary	Spelling	Mathematics
Hypothesis Mean Squared	1.2640	14.0004	8.0526
Univariate F	.5905	6.5183	2.7458
P Less Than	.4542	.0221	.1183

*F-ratio for multivariate test of context effect = 2.0703 with 3 and 13 degrees of freedom, p less than .1538.

SUMMARY TABLE FOR HYPOTHESIS 2; CONTEXT EFFECT ON THE ESTIMATES OF THE VARIANCE*

Computation	Vocabulary	Spelling	Mathematics
Hypothesis Mean Squared	2425.7690	7415.9271	352.3123
Univariate F	5.9923	14.4603	.3603
P Less Than	.0272	.0018	.5573

*F-ratio for multivariate test of context effect = 13.7555 with 3 and 13 degrees of freedom, p less than .0003.

SUMMARY TABLE FOR HYPOTHESIS 3; EXPOSURE EFFECT ON THE ESTIMATES OF THE MEAN*

Computation	Vocabulary	Spelling	Mathematics
Hypothesis Mean Squared	35.2272	56.0410	37.7817
Univariate F	25.4692	28.9180	16.4561
P Less Than	.0002	.0001	.0011

*F-ratio for multivariate test of exposure effect = 17.7319 with 3 and 13 degrees of freedom, p less than .0001.

SUMMARY TABLE FOR HYPOTHESIS 4; EXPOSURE EFFECT ON THE ESTIMATES OF THE VARIANCE*

Computation	Vocabulary	Spelling	Mathematics
Hypothesis Mean Squared	684.5130	691.2679	3.5178
Univariate F	.6222	1.0751	.0071
P Less Than	.4426	.3163	.9342

*F-ratio for multivariate test of exposure effect = .6625 with 3 and 13 degrees of freedom, p less than .5897.

TABLE V

DEVIATIONS OF MULTIPLE MATRIX SAMPLE ESTIMATES OF THE MEAN FROM ACTUAL POPULATION MEAN

ITBS* Subtest	Estimate 1	Estimate 2	Estimate 3	Estimate 4
Grade 4				
Vocabulary	- .582	.869	.101	1.022
Spelling	.491	-1.823	.009	-1.007
Mathematics	.307	.355	.452	1.662
Grade 5				
Vocabulary	- .350	1.667	.175	1.937
Spelling	1.037	-2.143	.310	- .193
Mathematics	- .067	.134	-1.025	1.546
Grade 6				
Vocabulary	- .632	.911	.360	1.021
Spelling	.668	-1.756	.447	- .654
Mathematics	- .957	1.129	- .251	2.170
Sum of Deviations	- .085	- .657	.578	7.504
Sum of Absolute Values of Deviations	5.091	10.787	3.130	11.212

*Iowa Tests of Basic Skills

TABLE VI

DEVIATIONS OF MULTIPLE MATRIX SAMPLE ESTIMATES OF THE VARIANCE FROM ACTUAL POPULATION VARIANCE

ITBS* Subtest	Estimate 1	Estimate 2	Estimate 3	Estimate 4
Grade 4				
Vocabulary	- .370	1.029	- 7.307	6.645
Spelling	-39.556	-13.876	17.232	-19.258
Mathematics	- 4.428	- 1.273	- 4.354	- 3.113
Grade 5				
Vocabulary	- .501	- 7.591	1.817	13.260
Spelling	-11.185	-13.700	-10.414	-20.676
Mathematics	2.223	6.119	2.537	- 2.364
Grade 6				
Vocabulary	- 1.432	- 7.294	- 4.045	22.760
Spelling	- 6.403	11.648	3.768	-25.184
Mathematics	16.044	3.274	- 6.233	12.576
Sum of Deviations	-45.608	-21.664	- 6.999	-15.354
Sum of Absolute Values of Deviations	82.142	65.804	57.707	125.836

*Iowa Tests of Basic Skills

In addition to the multivariate analysis of variance, deviation matrices were computed for both the estimates of the mean and the variance by subtracting the appropriate population parameter from each of the nine estimates (three subtests for each of three grade levels) for each of the four sets of matrix sampling estimates. Two summary indices were computed for each deviation matrix. The first, the sum of the deviations, was utilized as a measure of systematic bias. Estimates that were systematically too high would result in a large positive sum of the deviations, and estimates that were systematically too low would result in a large negative sum of the deviations. The second index, sum of the absolute values of the deviations, was an estimate of precision or variation. A relatively large sum of the absolute values of the deviations would indicate that the estimates varied considerably, while a relatively small sum would indicate that the estimates were relatively consistent.

The deviation matrices for the mean can be found in Table V (B) and the deviation matrix for variance in Table VI. An analysis of the deviation scores indicated that with the possible exception of Estimate 4, the sum of the deviation of the mean tended to sum to zero, i.e. there were not systematic differences. The multiple matrix estimates of the variance tended to be too low; however, the sum of the deviations again approached zero with the exception of Estimate 1. The large negative sum of deviations for Estimate 1 appears to be an artifact of a bizzare estimate for spelling at the fourth grade level. Multiple Matrix Sample Estimate 3, normal context-no previous exposure, was overall the most accurate set of estimates of both the means and variances. Estimate 4 tended to be the worst estimate.

On the basis of the analysis, the following conclusions were made:

1. The administration of the first set of matrix tests prior to the administration of the Iowa Tests of Basic Skills battery did not affect the examinee performance on the battery.

2. No evidence was found for the existence of a context effect in the multiple matrix sample estimates of the mean.

3. Change in item context significantly affected the estimates of the variance. The multiple matrix sample estimates of the variance computed from data collected by the actual administration of matrix tests showed greater variation on the deviation matrices than did the estimates computed from data collected during the administration of the entire battery. No evidence was found to indicate that the estimates of the variance were systematically larger or smaller than would have been expected.

4. Recent previous exposure to the items being sampled significantly affected the estimates of the mean. The estimates of the means computed from data that represented the examinees' second response to items within a week's time varied more than estimates computed from data that represented the examinees' first response to the items. Again, no evidence was found that the estimates of the mean were systematically larger or smaller than would have been expected.

5. Recent previous exposure to items in the sampled tests did not significantly affect the estimates of the variance. Both the means of the estimates and the variation about them were consistent for the four multiple matrix sample estimates.

Comparison of matrix sampling estimates with examinee sampling estimates. Multiple Matrix Sample Estimate 1 was the only matrix sample estimate used in this analysis. Estimate 1 approximated the way matrix sampling procedures would be used in an applied situation.

Deviation matrices and two summary statistics were computed for each of the ten sets of examinee sample estimates. The summary indices were used to identify the examinee sample estimates that when compared with Multiple Matrix Sample Estimate 1 would result in a conservative estimate of the precision of the matrix sample estimate. The deviation matrices for the means and variances are found in Tables VII and VIII and the two sets of summary statistics are found in Tables IX and X.

The set of examinee sample estimates that most accurately estimated the means and the set that most accurately estimated the variances were identified. A paired data *t test* was then used to compare the "most accurate" examinee sampling estimate with Multiple Matrix Sample Estimate 1.

The sum of the absolute values of the deviations for Multiple Matrix Sample Estimate (estimates of the means) was smaller than the sums of the absolute values of the deviations of all ten sets of the examinee sample estimates of the means. The paired data *t test* between Estimate 1 and the most accurate set of examinee sample estimates was significant in a direction favoring the multiple matrix sample estimates. Therefore, the multiple matrix sample estimates of the means were concluded to be significantly better than examinee sample estimates of the means.

TABLE VII

COMPARISON OF DEVIATIONS FROM THE POPULATION MEAN OF ESTIMATES OF THE MEAN OF MATRIX SAMPLE ESTIMATE 1 AND TEN EQUIVALENT RANDOMLY DRAWN EXAMINEE SAMPLING ESTIMATES

Estimate	Vocabulary	Spelling	Mathematics
Grade 4			
Matrix Sample	- .582	.491	.307
Examinee Sample 1	.442	-1.846	- .649
Examinee Sample 2	.005	- .131	-1.413
Examinee Sample 3	- .558	.179	- .649
Examinee Sample 4	.109	-2.988	- .887
Examinee Sample 5	- .177	.964	1.113
Examinee Sample 6	.680	.916	-1.268
Examinee Sample 7	1.442	1.440	.970
Examinee Sample 8	1.537	-1.322	-1.316
Examinee Sample 9	-3.225	1.773	- .887
Examinee Sample 10	.109	1.392	1.446
Grade 5			
Matrix Sample	- .350	1.037	- .067
Examinee Sample 1	-1.002	3.143	- .333
Examinee Sample 2	-1.202	1.082	-1.743
Examinee Sample 3	.348	-2.507	.967
Examinee Sample 4	-1.552	- .357	.767
Examinee Sample 5	-1.452	- .657	1.317
Examinee Sample 6	1.648	2.343	.217
Examinee Sample 7	- .102	- .207	- .083
Examinee Sample 8	-1.202	-1.157	- .283
Examinee Sample 9	1.148	.743	3.417
Examinee Sample 10	-1.702	1.443	.167
Grade 6			
Matrix Sample	- .632	.668	- .957
Examinee Sample 1	1.632	1.010	.717
Examinee Sample 2	-1.728	- .713	.500
Examinee Sample 3	2.743	- .102	- .728
Examinee Sample 4	-3.757	-2.157	- .172
Examinee Sample 5	.632	2.954	- .117
Examinee Sample 6	- .924	- .435	-1.950
Examinee Sample 7	-2.035	.343	- .894
Examinee Sample 8	- .035	-1.602	.161
Examinee Sample 9	- .035	-1.879	-2.561
Examinee Sample 10	- .979	-2.379	2.828

TABLE VIII

COMPARISON OF DEVIATIONS FROM THE POPULATION VARIANCE OF ESTIMATES OF THE VARIANCE OF MATRIX SAMPLE ESTIMATE 1 AND TEN EQUIVALENT RANDOMLY DRAWN EXAMINEE SAMPLING ESTIMATES

Estimate	Vocabulary	Spelling	Mathematics
Grade 4			
Matrix Sample	- .370	-39.556	- 4.428
Examinee Sample 1	6.274	28.108	3.689
Examinee Sample 2	- 3.490	1.498	4.072
Examinee Sample 3	- 5.026	- 7.392	6.589
Examinee Sample 4	15.040	- 1.778	- 5.377
Examinee Sample 5	5.526	- 8.049	14.523
Examinee Sample 6	-26.245	- .826	- 8.492
Examinee Sample 7	- 7.726	-10.649	- .134
Examinee Sample 8	-13.160	.622	-16.477
Examinee Sample 9	-24.426	3.217	23.523
Examinee Sample 10	-14.960	-28.678	- 4.277
Grade 5			
Matrix Sample	- .501	-11.185	2.223
Examinee Sample 1	14.114	.845	7.224
Examinee Sample 2	- 7.386	.850	-16.067
Examinee Sample 3	9.153	18.385	4.118
Examinee Sample 4	15.553	25.161	- 5.334
Examinee Sample 5	20.143	11.245	- 6.200
Examinee Sample 6	3.343	-11.597	- 9.526
Examinee Sample 7	- 2.592	15.363	- 9.632
Examinee Sample 8	- 7.844	22.613	-24.158
Examinee Sample 9	- 2.657	-16.102	- 5.421
Examinee Sample 10	13.261	37.519	10.013
Grade 6			
Matrix Sample	- 1.432	- 6.403	16.044
Examinee Sample 1	3.394	-14.487	12.797
Examinee Sample 2	24.304	- 9.014	- 6.197
Examinee Sample 3	.818	-22.017	5.098
Examinee Sample 4	-32.917	29.562	-13.961
Examinee Sample 5	-12.724	10.045	-10.879
Examinee Sample 6	27.642	.885	12.562
Examinee Sample 7	-22.646	65.709	- 6.382
Examinee Sample 8	43.119	- 4.517	- 6.039
Examinee Sample 9	3.707	-22.043	- 7.951
Examinee Sample 10	31.619	-17.131	14.510

TABLE IX

COMPARISON OF THE SUM OF DEVIATIONS AND THE SUM OF THE ABSOLUTE VALUES OF DEVIATIONS FROM THE POPULATION MEAN OF ESTIMATES OF THE MEAN OF MATRIX SAMPLE ESTIMATE 1 AND TEN EQUIVALENT RANDOMLY DRAWN EXAMINEE SAMPLING ESTIMATES

Estimate	Sum of Deviations	Sum of Absolute Values of Deviations
Matrix Sample	- .085	5.091
Examinee Sample 1	3.114	10.774
Examinee Sample 2	- 5.343	8.517
Examinee Sample 3	- .307	8.781
Examinee Sample 4	-10.994	12.746
Examinee Sample 5	4.577	9.383
Examinee Sample 6	1.227	10.381
Examinee Sample 7	.874	7.516
Examinee Sample 8	- 5.219	8.615
Examinee Sample 9	- 1.506	15.668
Examinee Sample 10	2.325	12.445

TABLE X

COMPARISON OF THE SUM OF DEVIATIONS AND THE SUM OF THE ABSOLUTE VALUES OF DEVIATIONS FROM THE POPULATION VARIANCE OF ESTIMATES OF THE VARIANCE OF MATRIX SAMPLE ESTIMATE 1 AND TEN EQUIVALENT RANDOMLY DRAWN EXAMINEE SAMPLING ESTIMATES

Estimate	Sum of Deviations	Sum of Absolute Values of Deviations
Matrix Sample	-45.608	82.142
Examinee Sample 1	61.958	90.932
Examinee Sample 2	-11.430	72.878
Examinee Sample 3	9.726	78.596
Examinee Sample 4	25.949	144.683
Examinee Sample 5	23.620	99.334
Examinee Sample 6	-12.254	101.118
Examinee Sample 7	21.311	140.833
Examinee Sample 8	- 5.841	138.549
Examinee Sample 9	-48.153	109.037
Examinee Sample 10	41.876	171.968

The sum of the absolute values of the deviations for Multiple Matrix Sample Estimate 1 (estimates of the variances) was smaller than the sums of the absolute values of the deviations of eight of the ten sets of examinee sample estimates of the variances. The paired data t test between Estimate 1 and the most accurate set of examinee sample estimates was not significant. Therefore, it was concluded that the multiple matrix sample estimates of the variances were as accurate as comparable examinee sample estimates of the variances.

Conclusions

This study once again demonstrated that multiple matrix sampling is an effective procedure for collecting data on the performance of groups. An a priori set of nine multiple matrix sample estimates, one for each of three subtests of the Iowa Tests of Basic Skills (Vocabulary, Spelling and Mathematics Concepts) for each of the three grade levels (fourth, fifth and sixth), was significantly more precise than ten similar sets of examine sampling estimates. No significant differences were found between the multiple matrix sample estimates and examinee sample estimates of the variances.

The findings regarding the effect of the changes in item context necessitated by matrix sample procedures and the effect of previous exposure to items on the matrix estimates were encouraging. The change in item context did not significantly affect the matrix sample estimates of the mean, but it did affect the estimates of the variance. Conversely, previous exposure to items affected the matrix sample estimates of the mean but not the estimates of the variance. Both the context and exposure effect involved an increase in the variation of the estimates and, therefore, a decrease in precision. Neither effect seemed to cause the estimates to be either systematically too high or too low. The loss in precision could be compensated for by increasing the number of observations. A systematic bias would have been much more vexing. The results, as encouraging as they were, should be interpreted cautiously. This study needs to be replicated in other settings using other instruments.